Published in final edited form as:

Lang Assess Q. 2015 ; 12(4): 386–408. doi:10.1080/15434303.2015.1100198.

Analysis of Bilingual Children's Performance on the English and Spanish Versions of the Woodcock-Muñoz Language Survey-R (WMLS-R)

Lia E. Sandilos,

University of Virginia, Charlottesville, Virginia

Kandia Lewis,

Nemours BrightStart!, Philadelphia, Pennsylvania

Eugene Komaroff,

Keiser University, Fort Lauderdale, Florida

Carol Scheffner Hammer,

Teachers College, Columbia University, New York, New York

Shelley E. Scarpino,

Bloomsburg University, Bloomsburg, Pennsylvania

Lisa Lopez,

University of South Florida, Tampa, Florida

Barbara Rodriguez, and

University of New Mexico, Albuquerque, New Mexico

Brian Goldstein

La Salle University, Philadelphia, Pennsylvania

Abstract

The purpose of this study was to investigate the way in which items on the Woodcock-Muñoz Language Survey Revised (WMLS-R) Spanish and English versions function for bilingual children from different ethnic subgroups who speak different dialects of Spanish. Using data from a sample of 324 bilingual Hispanic families and their children living on the United States mainland, differential item functioning (DIF) was conducted to determine if test items in English and Spanish functioned differently for Mexican, Cuban, and Puerto Rican bilingual children. Data on child and parent language characteristics and children's scores on Picture Vocabulary and Story Recall subtests in English and Spanish were collected. DIF was not detected for items on the Spanish subtests. Results revealed that some items on English subtests displayed statistically and practically significant DIF. The findings indicate that there are differences in the difficulty level of WMLS-R English-form test items depending on the examinees' ethnic subgroup membership. This outcome suggests that test developers need to be mindful of potential differences in

performance based on ethnic subgroup and dialect when developing standardized language assessments that may be administered to bilingual students.

Young bilingual students in the education system (11.2 million; NCES, 2009), particularly those of Spanish-speaking backgrounds, represent one of the fastest growing groups in the United States (Basterra, Trumbull, & Solano-Flores, 2010). Bilingual students currently make up approximately 30% of Head Start, federally funded U.S. preschool programming, with 80% of those children coming from Spanish-speaking homes and speaking a variety of Spanish dialects (Mathematica Policy Research, 2010). Because of the changing demographics of students in the United States, it is critical that linguistically and culturally appropriate tools are used for the assessment of Hispanic bilingual children (Basterra *et al.*, 2010). The goal of this study was to determine if items on tests of language proficiency function equally among Spanish-speaking ethnic dialect subgroups, as measured by the Woodcock-Muñoz Language Survey Revised Spanish and English versions (WMLS-R; Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005a, 2005b).

The importance of using linguistically and culturally valid assessments with bilingual children has been addressed through legislative policies (e.g., Individuals with Disabilities Education Act, 2004; No Child Left Behind Act of 2001). In addition, assessment guidelines have been established by a number of research and professional organizations, including the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME). The 2014 Standards for Educational and Psychological Testing advocated for high-quality assessment and reported that language-related factors, such as evaluating individuals with a test that is not in their first language, can significantly reduce the reliability and validity of scores (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Despite these efforts, improvements to existing assessment instruments are needed. In particular, language factors (e.g., monolingual vs. bilingual examinees, dialect differences) are not always sufficiently addressed in test development and assessment practices (Basterra *et al.*, 2010; Solano-Flores, 2006). Just as with monolingual individuals, tests for bilinguals need to be psychometrically strong, so that the scores function equally well for a wide variety of test takers.

Key Psychometric Considerations for Standardized Assessment of Bilinguals

It is critical that assessments for bilinguals are based on best practices for standardized test development and norming procedures (American Education Research Association *et al.*, 2014). Although a test may be published and accessible, this does not guarantee that it is linguistically valid for all populations (Barrueco, López, Ong, & Lozano, 2012; Peña, 2007). The norming and validation samples of assessments used with individuals of linguistically diverse backgrounds should demonstrate representativeness of the population demographics (race, socioeconomic status, age, geographic region, etc.), and the test should also report separate parallel examinations of data for members of different linguistic subgroups to determine the validity of use with those groups (American Education Research *et al.*, 2014).

For example, if a Spanish-language test is intended for use with Spanish-English bilingual children living in the United States, particular attention should be paid to the inclusion of Spanish-English bilingual children living on the United States mainland who speak a variety of dialects during the standardization process. For the current study, Picture Vocabulary and Story Recall subtests of the WMLS-R were examined by using differential item functioning to determine if test items function equally well for ethnic subgroups of Hispanic bilingual children who speak different dialects of Spanish and reside in the United States.

Development of Woodcock-Muñoz Language Survey—Revised (WMLS-R)

The WMLS-R Spanish and English versions are composed of subtests focused on academic language acquisition (Alvarado, Ruef, & Schrank, 2005; Cummins, 1979). Since its initial publication, the WMLS has been widely used in research and practice to evaluate children's proficiency in both languages (Hammer *et al.*, 2012; Mahon, 2006). In the school setting the results of WMLS-R language proficiency assessment helps educators to determine which children may benefit from placement in English as a second language (ESL) classrooms and which children may have a language disorder that will necessitate further assessment and may also require special education services (Alvarado *et al.*, 2005). Thus, it is critical that scores on this measure are valid for bilinguals.

Although the WMLS-R Spanish- and English-form standardization samples included a large number of participants with diverse characteristics, variations in performance across the Hispanic bilinguals representative of major ethnic subgroups and dialects in the United States was not assessed as part of the Spanish or English standardization process. The WMLS-R Spanish version was standardized on a subsample of 1,157 monolingual Spanish-speaking children from several Spanish-speaking countries, as well as Spanish-speaking children in the United States. Specifically, the standardization manual characterizes the Hispanic sample used for the Spanish version as “native Spanish speakers from outside the United States, and monolingual or near-monolingual Spanish speakers from within the United States” (Alvarado *et al.*, 2005, p. 78). Large numbers of bilingual children were not included in the sample. For the inclusion of ethnic subgroups, approximately a quarter (26%) of the sample was of Mexican descent; however, there were very few Puerto Rican participants (2%) and no Cuban participants included in the sample, though these combined groups comprise 13% of the U.S. Spanish-speaking population (Alvarado *et al.*, 2005). In addition, Spanish-speaking participants living in the United States comprised a very small percentage of the full standardization sample (7%). The underrepresentation of Spanish-speaking participants who are illustrative of U.S. Hispanic bilinguals in the normative sample raises some questions about the measure's ability to accurately assess language abilities among bilingual examinees learning Spanish and English simultaneously in the United States.

Normative data for the WMLS-R English version were collected from nearly 9,000 English-speaking participants from over 100 counties in the Northeastern, Midwestern, Southern, and Western regions of the United States (Alvarado *et al.*, 2005). Stratified random sampling controlled for both individual (e.g., race, sex, education level) and community (e.g., census region, community size) characteristics. Although the manual reports that the

standardization sample of the English version excluded “subjects who had less than 1 year of experience in an English-speaking classroom” (Alvarado *et al.*, 2005, p. 75), there is no indication of the number of bilingual participants included from each region or ethnic subgroup in the United States. The lack of normative information for bilingual participants included in the WMLS-R English sample is a limitation of the measure and warrants further investigation with Hispanic ethnic subgroups. A statistical method that can be used to explore differences in test performance as a result of membership in a sociodemographic group is differential item functioning.

Differential Item Functioning

For test scores to be used accurately and meaningfully with examinees of varying characteristics, the test content must demonstrate equivalence or “test fairness” across test takers. The examination of individual characteristics that may affect test performance is the primary goal of differential item functioning (DIF) analysis. DIF analysis is often used to develop new tests and can also be used to adapt existing measures to new settings, languages, and/or cultures (Zumbo, 2007). DIF occurs when more test takers in one group answer a test item correctly compared to an equal number of knowledgeable individuals in another group (Magis, Béland, Tuerlinckx, & De Boeck, 2010). An item is said to function differently, or exhibit “test bias,” when students with the same ability level, but from different groups, have significantly different probabilities of answering the item correctly.

The WMLS-R standardization manual (Alvarado *et al.*, 2005) does not report the use of differential item functioning analysis of Spanish-English bilinguals who speak varying Spanish dialects. Moreover, data examining the differences in performance on the WMLS-R across bilingual individuals of varying Spanish ethnic-dialect subgroups have not been reported in published literature to date. This study addresses this concern by examining the influence of ethnic-dialect subgroup membership on test performance.

The Role of Ethnicity and Dialect

Within the Hispanic population in the United States, the three largest ethnic subgroups are Mexican, Cuban, and Puerto Rican. Individuals within each of these ethnic subgroups have a shared sense of membership as well as shared cultural traditions, beliefs, and values that make them unique from other subgroups (Wolfram, 1991). A key characteristic of each ethnic subgroup is the dialect of Spanish and English that is spoken. Dialect is defined as “any language variety that typifies a group of speakers within a language” (Wolfram, 1991, p. 2). The dialects spoken by each group are influenced by the family country of origin, the geographic region of the United States in which they reside, as well as their social and economic background (Basterra *et al.*, 2010; Silva-Corvalán, 2004; Wolfram, Carter, & Moriello, 2004). As a result, there are differences among the Mexican, Cuban, and Puerto Rican dialects of Spanish in pronunciation, grammar, and vocabulary (Silva-Corvalán, 2004; Wolfram, 1991; Zentella, 2004). Many basic words can take different lexical forms or meanings, and words can vary in frequency of usage depending on the dialect spoken. These differences may affect the performance of children from these three ethnic groups on the Spanish subtests of the WMLS-R. The authors of this study who include native Spanish

speakers examined WMLS-R items and identified several dialectal differences across Mexican, Cuban, and Puerto Rican dialects. For example, “baby” may be referred to as *niño*, *chiquito*, and *infante* in Mexican, *bebé* in Cuban, and *nené* in Puerto Rican, “pencil” could be called a *lápiz* in Mexican or *lapicero* in Cuban and Puerto Rican, and “eyeglasses” may be most commonly called *anteojos*, but could also be referred to as *antiparras* or *gafas* in Mexican.

In addition to lexical variations, there are syntactic and phonological differences that can occur across dialects. One example of a dialectal difference between the Spanish subgroups related to syntax is the preference of Puerto Rican speakers to use the preverbal position of the subject pronoun (Qué tú dices? “What are you saying?”), similar to the English language, whereas Cuban and Mexican speakers will use the postverbal position (Qué dices tú?; Santoro, 2007). For morphosyntax, the deletion of the final /s/ in a word occurs in both the Puerto Rican and Cuban dialects but not in the Mexican dialect (e.g., día vs. días). An example of a variation in phonology is the elision of /x/ that occurs primarily in the Mexican dialect such that *trabajo* (job) would be pronounced *traba-o* (Michnowicz, 2012).

Differences in children’s performance on the English subtests of the WMLS-R may also be observed between these three ethnic subgroups because of the variations in the dialects of English that they speak. These differences may be the result of at least three factors. The first factor is the influence of the first language on the second. “For ethnic groups that maintain a language other than English, there is the potential of language transfer from the other language which is ‘fossilized,’ stabilized and perpetuated as part of the English variety used by members of the ethnic group” (Wolfram, 1991, p. 104). The second factor stems from the generalized language learning strategies that groups may use when acquiring English. Thus, these first two factors impact children’s performance on the WMLS-R, and in particular on the Story Recall subtest. The third factor is the dialect of English that is spoken by the monolinguals and the bilinguals in the region of the United States in which they reside (Arrieta, 1994; Barrón & San Roman, 2013).

Additional Sources of Variation in Language Proficiency Outcomes

Although the children’s ethnic-dialect subgroup is the primary focus of this investigation, it should be acknowledged that additional sociodemographic and contextual factors can influence children’s language performance. For example, the amount of exposure children have to Spanish and English has consistently been shown to affect their abilities in two languages, with increased exposure relating to stronger skills (Dunn Davison & Hammer, 2012; Hammer *et al.*, 2012; Place & Hoff, 2011). The length of time children have lived in the United States and the geographic region in which they reside can influence language exposure. Children living in the United States for longer periods of time often have more exposure to English (Arcia, Skinner, Bailey, & Correa, 2001; Cabrera, Shannon, West, & Brooks-Gunn, 2006; Hammer *et al.*, 2012). However, in regions where large groups of Spanish speakers live, children may actually have increased exposure to Spanish and less exposure to English. For instance, New Mexico has the one of the largest concentrations of Mexican Spanish speakers in the United States (Silva-Corvalán, 2004), and the National Center for Education Statistics has documented that the percentage of children who have

difficulty speaking English is significantly higher for children of Mexican descent than for children of Puerto Rican or Cuban descent (Aud, Fox, & KewalRamani, 2010).

For children's vocabulary development, the context in which children are exposed to various vocabulary items influences whether children learn words in Spanish or English. Specifically, bilingual children tend to learn nonacademic vocabulary words in their home language and academic words in the language of schooling (Schleppegrell, 2012), particularly when they live in homes and communities where Spanish is used most of the time. Specific to this study, the majority of the initial items on the Spanish and English Picture Vocabulary subtests of the WMLS-R are related to items that may be learned at home. Examples include English words, such as window, ball, mouth, cat, house, chair, sock, and apple, as well as Spanish words including globos (ball), árbol (tree), ojo (eye), casa (house), and manzana (apple).

Finally, socioeconomic status, often measured by maternal education, has also been shown to affect children's language abilities (Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010; Golberg, Paradis, & Crago, 2008; Hammer *et al.*, 2012). Previous research has identified a positive relationship between maternal education level and bilingual children's English vocabulary size (Goldberg *et al.*, 2008; Quiroz, Snow, & Zhao, 2010). Hammer and colleagues (2012) found that bilingual children whose mothers had completed higher levels of education performed better on measures of vocabulary and story recall in their two languages.

Given that background characteristics and contextual factors can influence children's language proficiency in English and Spanish, it is important that these factors are taken into consideration when examining how bilingual children perform on language proficiency tests (Barrueco *et al.*, 2012). Thus, various sociodemographic and language usage/exposure covariates were also controlled for in the differential item functioning analyses.

Purpose of the Investigation

Given the large number of bilingual Spanish-English speaking students in the United States and the variety of factors that can contribute to language proficiency, it is important to understand the way in which items on standardized tests function for diverse Spanish-English bilinguals. Examination of the norming procedures used to develop the Spanish and English versions of the WMLS-R indicates that bilinguals from the United States were not well represented in the standardization sample of both the Spanish and English versions and that ethnic subgroup differences in performance have not been explored. As such, the purpose of the present study was to investigate the way in which Spanish and English items on the WMLS-R varied based on Spanish ethnic-dialect subgroup. Differential item functioning was used to assess Picture Vocabulary and Story Recall subtests of the WMLS-R in both Spanish and English. Within each subtest, items were examined to determine whether they functioned similarly for bilingual children of Mexican, Puerto Rican, and Cuban backgrounds. For each of the four subtests, it was hypothesized that items would function differently depending on membership to an ethnic-dialect subgroup.

METHOD

Participants

The participants consisted of bilingual preschool and kindergarten children who participated in a larger study of bilingual children's phonological development ($N = 448$). Participants were recruited from Head Start programs, community-based preschool programs, and school districts. The participants resided in urban areas of central New Mexico, southeastern and central Florida, and central Pennsylvania.

To be included in this study, children's mothers had to report speaking to their children in a Cuban, Puerto Rican, or Mexican dialect of Spanish from birth, and children had to have data on all outcomes and covariates of interest ($N = 250$). The children had no cognitive, physical, or neurological impairments and no concerns about their speech and language abilities or hearing as reported by the children's mothers. Across the dialect subgroups, children's average age ranged from 58.32 to 60.49 months (see Table 1).

The Spanish dialects spoken in the home consisted of 42% Mexican, 35.2% Cuban, and 22.8% Puerto Rican. The majority (97%) of children speaking the Cuban dialect lived in Florida, 83% of children speaking the Mexican dialect lived in New Mexico, and 51% of Puerto Rican speakers lived in Pennsylvania while the other 49% lived in Florida. Across the three subgroups, more than half of mothers (57.89%–88.57%) and fathers (69.32%–78.10%) reported using primarily “more or all” Spanish with their children, and children were exposed to more English at school than at home (see Tables 2 & 3).

Measures

Woodcock Muñoz Language Survey—Revised (WMLS-R)—The WMLS-R is a standardized test of language proficiency that can be administered to examinees from age 2 to 90. Two subtests of the WMLS-R Spanish and English versions, Picture Vocabulary/ Vocabulario Sobre Dibujos and Story Recall/ Rememoración de Cuentos, were examined in this study. It should be noted that the English and Spanish versions of the subtests contained different items.

The first subtest, Picture Vocabulary, assesses children's vocabulary knowledge. The subtest was selected for this study because vocabulary generally reflects overall language abilities and has been significantly related to early literacy and later reading abilities (August, Carlo, Dressler, & Snow, 2005). Vocabulary knowledge also is often used as a measure of language proficiency in bilingual children. English Picture Vocabulary consists of 59 items and Spanish Picture Vocabulary consists of 58 items. The first two questions on the English subtest are receptive items, and the first six questions on the Spanish subtest are receptive items. For each receptive-vocabulary item, children were asked to point to a picture named by the examiner within a group of pictures. The remaining expressive items on the subtests require the children to name each picture with items increasing in difficulty. The median internal consistency reliability coefficient for Picture Vocabulary is .91.

On the second subtest, Spanish and English Story Recall, expressive language, language comprehension, and memory are assessed, which, combined with the vocabulary subtest,

provide a snapshot of bilingual children's language skills. The Spanish and English subtests are both composed of 11 stories that are read to the children. Each story consists of essential information that children are asked to recall, and children receive one point for each piece of information they include when retelling the story. The length of each story ranges from one sentence to several sentences. The median internal consistency reliability coefficient for scores on this subtest is .76.

Procedures

Data collectors were formally trained in all measures used in the study. Demographic and language data were collected from a questionnaire, which was administered to children's mothers in person or over the phone. Data collectors read each of the questions and recorded the mothers' responses into a computer using a software program known as Study Participant (Knightsoft, 2008). The questionnaires took 25–30 minutes to complete.

Each child was administered the four subtests of the WMLS-R in Spanish and English by trained bilingual data collectors. Different data collectors conducted separate testing sessions at least one week apart in each of the languages. Each testing session lasted approximately 15–20 minutes. The language in which the children were first tested was counterbalanced.

Data Analysis

This study investigated DIF across three Spanish subgroups: Mexican, Cuban, and Puerto Rican. DIF analyses were used to examine dichotomous test items from English Picture Vocabulary, Spanish Picture Vocabulary, English Story Recall, and Spanish Story Recall subtests of the WMLS-R (Alvarado *et al.*, 2005). SAS (v. 9.3, SAS Institute Inc., Cary, NC) was used for all analyses. Variables were screened for potential anomalies using frequency, means, and graphics procedures. All subsequent inferential analyses involving DIF were conducted with the generalized mixed model procedure called PROC GLIMMIX. Specialized IRT software, such as WINSTEPS, was not used in this case because these programs do not easily accommodate covariate adjustment for DIF analysis.

The covariates selected in this analysis were shown in the literature to affect language outcomes of bilingual children (Bohman *et al.*, 2010; Duursma, Romero-Contreras, Szuber, Proctor, & Snow, 2007; Hammer, Davison, Lawrence, & Miccio, 2009; Hammer *et al.*, 2012). The covariates consisted of children's age in months, length of time children have lived on the U.S. mainland (1 = *since birth*, 2 = *not since birth*), maternal education level (1 = *less than high school*, 2 = *high school*, 3 = *some college/trade school*, 4 = *college and beyond*), and language(s) spoken to children by their mother, father, and teacher, as well as language(s) spoken by children to their mother, father, and teacher (1 = *All Spanish*, 2 = *More Spanish than English*, 3 = *Equal Spanish and English*, 4 = *More English than Spanish*, 5 = *All English*). Although the direct impact of these variables was not examined, they were held constant at their overall means when testing for post hoc pairwise differences in difficulty between the three subgroups.

Mixed logistic regression models were used for the DIF analysis with persons as random effects and items as fixed effects (Van den Noortgate & De Boeck, 2005), and models

controlled for sociodemographic and language usage/exposure covariates that potentially could have been the source for DIF.¹ This statistical methodology is similar to the hierarchical linear model (HLM) approach for DIF analysis (Kamata, 2001). The current analysis used one ordinal variable, instead of many dummy variables, to parameterize item difficulty. Because the WMLS-R test items become progressively harder within each subsection or item, difficulty was captured by the beta coefficient on the ordinal variable. For every unit change (next item on the ordinal scale), the beta coefficient is an estimate for the change in the logit of getting the item wrong. The three subgroups (Cuban, Puerto Rican, and Mexican) were coded as categorical variables. A statistically significant interaction ($p < .05$) between the ordinal “item” variable and the categorical ethnic subgroup variable, with other covariates held constant, indicated that item difficulty was not equal for all three groups.

The significant interaction was followed by post hoc pairwise analyses to determine which groups differed and on what item(s). Specifically, three post hoc comparisons (Cuban vs. Mexican, Cuban vs. Puerto Rican, and Mexican vs. Puerto Rican) with LSMEANS were run at each item number along with a Bonferroni adjustment² for the p -values. The p -values from pairwise contrasts were submitted to another overall Bonferroni adjustment with an SAS procedure called PROC MULTTEST to guard against additional alpha inflation. Finally, statistical significance of differences in difficulty were identified (overall Bonferroni p -value $< .05$), but substantive significance (i.e., differences in difficulty exceeding $|0.5|$ on the logit scale; Smith & Smith, 2004; WINSTEPS) was also required before items were considered as demonstrating DIF.

RESULTS

The final subsamples examined for each WMLS-R subtest were determined by listwise deletion, so that participants were required to have all data on the four outcome variables and nine covariates to be included in the study. The final subsamples were as follows: English Picture Vocabulary and Story Recall ($N = 241$), Spanish Picture Vocabulary ($N = 250$), Spanish Story Recall ($N = 249$). Cross-tabulations and paired t tests were then used to compare test scores and covariates across the final samples. The results indicated that the subsamples were not significantly different statistically. All sociodemographic and language usage/exposure covariates are reported for each ethnic subgroup within the largest subsample ($N = 250$; Tables 1–3).

Statistically significant differences ($p < .05$) were identified among three dialect subgroups across covariates, as indicated by subscripts in Tables 1–3. Mothers of the Mexican subgroup reported significantly less education than the mothers of the Puerto Rican and Cuban subgroups. Specifically, 84.7% of Mexican mother's did not receive a GED or high school diploma, compared to 44.4% of Cuban mothers and 59.7% of Puerto Rican mothers

¹DIF was examined without covariates, as well as with geographic region of inhabitation included as a covariate, and the same findings were produced.

²Bonferroni adjusted p -values were obtained by multiplying raw p -values by three (the number of post hoc comparison tests). Next, the initial adjusted p -values received a second Bonferroni adjustment, in which the p -value was multiplied by the number of items assessed within the subtest (Westfall, Tobias, & Wolfinger, 2011).

(Table 1). In addition, significantly more Mexican (86.7%) and Cuban children (87.5%) were born on the U.S. mainland than Puerto Rican children (56.10%). Significant differences were also identified across subgroups for language usage and exposure covariates. For example, Puerto Rican children were exposed to significantly more English from their fathers (21.05%) than Mexican children (7.62%), and both Puerto Rican (50.88%) and Cuban (22.73%) children used significantly more English with their mothers than Mexican children (3.81%) (see Table 3).

Raw and standard score means and standard deviations for Spanish and English Story Recall and Picture Vocabulary subtests for each dialect are reported in Table 4. Analyses were conducted on raw scores; standard scores were included to assist the reader with interpretation of children's performance on the subtests (see Table 4). An examination of the skewness and kurtosis values indicated that test scores approximated a normal range and a visual inspection of probability plots indicated linearity of the scores (Field, 2009).

All analyses modeled the probability of receiving an incorrect score on an item. Figure 1 displays the probability of children from each subgroup providing an incorrect response. Figure 2 displays the predicted logit of incorrect responses. The predicted logit is a transformation of probability in which numbers increasing in value convert into a greater probability of getting an item wrong. As such, larger negative values and then increasingly positive values indicate increased probability of responding incorrectly. If subgroups exhibit statistically significantly different predicted probabilities, this is an indication that the item is functioning differentially across the groups.

Spanish WMLS-R

DIF was examined for all items (1–58) on the Spanish Picture Vocabulary subtest scores. No statistically or substantively significant DIF was detected for any items on the subtest. Figure 1 displays the probability of children from each subgroup providing an incorrect response. The three dialect groups did not demonstrate significantly different predicted probabilities of receiving an incorrect score for each Spanish picture vocabulary item. Figure 2 displays the predicted logit of incorrect responses. None of the logit values for each group were statistically significantly different indicating again that the groups had similar likelihood of getting an item wrong. In addition, the Spanish Story Recall subtest scores were assessed, and again no statistically or substantively significant DIF was detected on the first seven passages of Spanish Story Recall subtest (Figures 3 and 4).

English WMLS-R—Scores on Picture Vocabulary and Story Recall subtests of the WMLS-R English version were also inspected for DIF. The English Picture Vocabulary subtest items of the WMLS-R showed statistically and substantively significant DIF between the Mexican and Cuban subgroups. Specifically, the interaction between item difficulty and subgroup was statistically significant after accounting for the covariates. Significant covariates in the analyses included the language(s) spoken by children to their teachers ($\beta = -.70, p < .0001$), children's age ($\beta = -.10, p < .0001$), the amount of time children have not lived on the U.S. mainland ($\beta = 1.18, p = .001$), and maternal education level ($\beta = -.57, p < .0001$). A negative beta weight represents the decrease in the probability

of getting an item wrong with every unit increase in the value of the covariate. Conversely, a positive beta weight represents the increase in the probability of getting an item wrong with every unit increase in the value of the covariate. The probability of getting an English item wrong decreased for every unit increase in children's language usage with their teacher, children's age, and mothers' education level, whereas the probability of getting an item wrong increased for every unit increase in the amount of time children spent not living on the U.S. mainland.

Of the 59 items on the English Picture Vocabulary subtest, DIF was examined for items 1–36, at which point all participants reached the subtest ceiling (i.e., the point at which all students in the sample received enough incorrect responses to discontinue the subtest). Analyses revealed that statistically and practically significant DIF was operating on the first 14 test items between Mexican children and Cuban children. In other words, the children who spoke a Mexican dialect had a higher likelihood of answering items 1–14 incorrectly than did the children who spoke a Cuban dialect (Figures 5 and 6). As shown in Figure 5, the nonlinear regression lines for predicted probabilities exhibit some visual separation between groups, which is highlighted by a circle. Figure 6 displays the same differences in terms of logits, which are linear. For the initial Picture Vocabulary items, the Mexican subgroup displayed smaller negative logit values than the Cuban and Puerto Rican subgroups indicating a greater likelihood of getting an item wrong. This difference was found to be statistically significant between the Mexican and Cuban subgroups. These test items, their logit-difference values, and *p*-values are reported in Table 5.

On the English Story Recall subtest, statistically and practically significant DIF was detected among all three subgroups. The language(s) spoken by children to their teachers ($\beta = -.31, p = .01$), children's age ($\beta = -.06, p < .0001$), and maternal education level ($\beta = -.49, p < .0001$) were all statistically significant covariates. In other words, the probability of getting an item wrong decreased for every unit increase in children's language usage with their teacher, children's age, and mothers' education level. DIF was examined for story passages 1–5 containing items 1–27, but story passages 6–11 were not assessed because all participants reached the subtest ceiling at story passage 5. Statistically and substantively significant DIF was identified on story passage 5 (items 18–27) between Mexican and Puerto Rican children indicating that Mexican children had a higher likelihood of responding incorrectly than the Puerto Rican children. There was also significant DIF between Cuban and Puerto Rican children on the last two items (26 and 27) of English Story Recall passage 5, where Cuban children had a higher likelihood of responding incorrectly than children who spoke a Puerto Rican dialect. As shown in Figure 7, the probability of getting an item wrong increased with item difficulty (increasing item number) for all three groups; however, there was a distinct separation among items 18–27, with Mexican and Cuban children exhibiting higher positive logits than Puerto Rican children (Figure 8). In other words, these items were easier for Puerto Rican than Mexican or Cuban children. Test items, logit-difference values, and *p*-values are reported in Table 6. Overall, across the two English subtests, Mexican children consistently demonstrated a higher likelihood of answering DIF items incorrectly than Puerto Rican or Cuban children.

DISCUSSION

The purpose of this study was to investigate the ways in which children's performance on Spanish and English items of the WMLS-R Picture Vocabulary and Story Recall subtests varied on the basis of Spanish-dialect subgroup. Differential item functioning (DIF) was used to analyze the WMLS-R test items, after controlling for sociodemographic and language variables within the analyses. DIF was not detected for items on the Spanish Picture Vocabulary and Story Recall subtests. Results revealed that some items on English Picture Vocabulary and Story Recall subtests displayed statistically and practically significant DIF. In other words, children in the Mexican subgroup demonstrated a higher likelihood of providing an incorrect answer than Cuban children on the first 14 items of English Picture Vocabulary. On one of the five English Story Recall passages analyzed, children in the Mexican and Cuban subgroups demonstrated a higher likelihood of providing an incorrect answer than Puerto Rican children.

The lack of DIF on the WMLS-R Spanish version is an important and positive outcome indicating that test items within the two subtests function equally well for Spanish-speaking bilinguals of Mexican, Cuban, and Puerto Rican backgrounds. This finding indicates that WMLS-R Spanish items are appropriate for young bilingual students living in the United States, who are members of these three ethnic subgroups, despite the underrepresentation of Spanish-speaking participants reflective of U.S. Hispanic bilinguals in the normative sample. Thus, this study lends support for the generalizability of the Picture Vocabulary and Story Recall subtest scores of the WMLS-R Spanish version across subgroups of Spanish-English bilingual examinees.

Conversely, DIF was present across the three Spanish subgroups for the English WMLS-R subtests. These findings indicate that the items may be biased for certain Spanish dialectal groups on the English version of this language proficiency measure. In particular, the issue of item bias was most salient for Mexican children. A limited number of bilingual participants included in the English WMLS-R normative sample is one possible contributor to the presence of DIF among the subgroups, though it is difficult to explain why a suboptimal normative sample might be an issue for the English version, but not for the Spanish version. Specific reasons for item bias or differences in performance among Mexican, Cuban, and Puerto Rican children on the English subtests are difficult to identify; however, one could speculate that factors such as ethnicity-dialect, language exposure, geographic region of inhabitation, and socioeconomic status (SES) are variables potentially contributing to these findings. One possible explanation for the presence of Spanish subgroup differences in the English Picture Vocabulary and Story Recall subtests could be the relation among ethnicity, dialect, and language exposure and usage. Ethnicity and dialect are intertwined, and young children initially learn to communicate in the context of their home language, which is influenced by the surrounding ethnic and linguistic community (Parlakian & Sánchez, 2006). Ethnic community influences the activities and conversations that occur in the home environment, which in turn impact the frequency of word usage in both first and second languages. Therefore, it may be that differences in the linguistic practices of the subgroups affected children's English vocabulary knowledge.

Previous research has shown that bilingual children typically learn certain vocabulary words in specific languages and settings (Bedore, Peña, García, & Cortez, 2005; Umbel, Pearson, Fernandez, & Oller, 1992), with nonacademic/informal, high-frequency words learned at home in the native language and academic, subject-specific vocabulary learned in the school setting in English (Schleppegrell, 2012). Nearly all items completed on the Spanish and English Picture Vocabulary subtests contained basic nonacademic vocabulary (e.g., *ball*, *cat*, *house*); thus, it is possible that this young sample of children was less familiar with the test items in English because they primarily hear and use those words in Spanish. Although most parents reported using more Spanish than English with their children, mother-report indicated that the Mexican children not only had more exposure to Spanish in the home but they also had more Spanish exposure in the school, indicating that the sociocultural environment of the Mexican children was highly supportive of the Spanish language. In contrast, the Puerto Rican and Cuban samples reported higher levels of English usage and exposure than Mexicans in the home and school environments, with the Puerto Ricans reporting the most English. Because Spanish was primarily used in the home and used frequently in school, children in the Mexican subgroup may not have been exposed as much to the English words and constructions that were tested by the WMLS-R. Moreover, the frequency and usage of vocabulary words that are Spanish-English cognates can differ by subgroup, resulting in English items that may be more sensitive to DIF. For example, in Cuban Spanish, *boat* is often referred to as its cognate, *bote*, but may be referred to as *barco* in other dialects, and *cone* may be most commonly referred to as its cognate, *cono*, in Cuban, but may be called *cucurucho* when related to *ice cream* in other dialects.

Limited exposure to the English language also may have played a role in Mexican children's performance on the Story Recall subtest, as limited familiarity with English sentence structure could make the task more taxing on children's working memory and more difficult for children to understand and recall the story. For example, within the Story Recall subtest, items in story passage 5 exhibited DIF. This passage was longer and more syntactically complex than the prior passages. Some examples of structural differences between English and Spanish in passage 5 were identified. The first example involves a difference in word order. In English, the adjective comes before the noun (*The baby bear*); however, adjectives come after the noun in Spanish (*El oso jovo*). The second example relates to the use of particles. A verb paired with different particles in English may result in a completely different verb choice in Spanish. For example, in English the verb "climbing" can be paired with a different particle (up/down) to change meaning (climbing up vs. climbing down) and in Spanish the verb used to express these meanings may be completely different (*subiendo* vs. *bajando*). Although selected background variables were accounted for in this study, the relation among ethnic-dialect subgroup and language exposure and usage is complex and may not have been fully captured through the language items on the background questionnaire.

It is also important to recognize that the geographic region of language use could influence both the Spanish and English dialects of the speaker (Goldstein & Iglesias, 2001; Goldstein, 2001). Region and ethnic-dialect subgroup are difficult constructs to separate as ethnic groups typically settle together. In the current study the subgroups exhibiting DIF each

resided in primarily one location, with Cubans living in Florida and Mexicans residing in New Mexico. The Puerto Rican sample was less concentrated in one area, because children resided in both Pennsylvania and Florida. The state of New Mexico has one of the largest concentrations of Mexican-Spanish speakers in the United States (Silva-Corvalán, 2004), and national education data have indicated that children of Mexican background have more difficulty speaking English than those of Puerto Rican or Cuban backgrounds (Aud *et al.*, 2010).

Bilingual children have also demonstrated different response patterns based on their geographic location and their particular dialect (Arrieta, 1994). An example of this is “Tejano/Chicano English,” often spoken by Mexican Americans living in the Southwestern United States, in which the speakers can differ from Standard English in phoneme production, intonation, and frequency of vocabulary usage (Arrieta, 1994; Barrón & San Roman, 2013). Additional research examining variations in English production by bilinguals living in different geographic regions of the United States would help to further understand the potential interaction of Spanish ethnic-dialect subgroup, geographic location, and English language production.

Although maternal education was controlled for within the analyses as a measure of SES, reported yearly income may be a more accurate conceptualization of financial status because immigrant families who were educated in their home country may have more difficulty finding work in the U.S. mainland. Fuligni and Yoshikawa (2003) reported that the SES of immigrant families is a much more complex variable than that of non-immigrant families because some immigrants may be higher SES in their home country but may drop in SES in the United States because of language barriers, difficulty obtaining work, etc. Data for the education and SES of Hispanics residing in the United States have revealed that differences exist among Hispanic ethnic subgroups. In 2010 Mexicans demonstrated higher poverty rates than Cubans or Puerto Ricans residing in the United States (Motel & Patten, 2012). On the basis of these poverty rates, it may be that the Cuban and Puerto Rican children had more access to English-language resources in the home (e.g., books or television) than the Mexican children, resulting in differences in performance on English literacy tasks. Future studies should examine both education level and annual income as indicators of SES.

The findings of the present study support the need for conducting separate analyses on ethnic subgroups to determine the tests’ generalization across those groups. Future studies of DIF within test scores of bilingual populations may wish to also conduct a detailed error analysis to identify potential patterns of incorrect responses among subgroups. Further investigation of DIF between bilingual and monolingual examinees on the Spanish and English versions of the WMLS-R is also warranted.

Overall, the Spanish language subtests of the WMLS-R did not exhibit DIF. This positive finding supports the use of the Spanish Picture Vocabulary and Story Recall subtests as measures of Spanish language proficiency that function equally well for Mexican, Cuban, and Puerto Rican bilingual examinees. However, caution should be taken when administering the English version to Spanish-speaking bilinguals, particularly those of Mexican descent, because the test did not function equally well for these subgroups.

Because test bias could lead to an inaccurate diagnosis of a language disorder (Peña & Halle, 2011), test developers need to be mindful of potential differences in performance based on ethnic subgroups and of the importance of including bilingual participants in standardization samples when developing and norming language assessments. In addition, when selecting tests for bilingual examinees, educators should assess students in both languages and take time to consider measures that are most appropriate for their needs based on the measure's content and validation procedures.

REFERENCES

- Alvarado, CG.; Ruef, ML.; Schrank, FA. Woodcock-Muñoz Language Survey-Revised comprehensive manual. Itasca, IL: Riverside Publishing; 2005.
- American Education Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
- Arcia E, Skinner M, Bailey D, Correa V. Models of acculturation and health behaviors among Latino immigrants to the US. *Social Science & Medicine*. 2001; 53(1):41–53. [PubMed: 11386307]
- Arrieta, O. Language and culture among Hispanics in the United States. In: Weaver, T., editor. *Handbook of Hispanic cultures in the United States: Anthropology*. Houston, TX: Arte Publico Press; 1994. p. 168-186.
- Aud, S.; Fox, M.; KewalRamani, A. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office; 2010. Status and trends in the education of racial and ethnic groups (NCES 2010-015).
- August D, Carlo M, Dressler C, Snow C. The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice*. 2005; 20(1):50–57.
- Barrón CC, San Roman J. Teachers guide to supporting Mexican American standard English learners. Academic English Mastery Program. 2013 Retrieved from <http://home.lausd.net/index.jsp>.
- Barrueco, S.; López, M.; Ong, C.; Lozano, P. Assessing Spanish-English bilingual preschoolers. Baltimore, MD: Paul Brookes Publishing Co; 2012.
- Basterra, MR.; Trumbull, E.; Solano-Flores, G. Cultural validity in assessment: Address linguistic and cultural diversity. New York, NY: Routledge; 2010.
- Bedore LM, Peña ED, García M, Cortez C. Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools*. 2005; 36(3):188–200.
- Bohman TM, Bedore LM, Peña ED, Mendez-Perez A, Gillam RB. What you hear and what you say: Language performance in Spanish-English bilinguals. *International Journal of Bilingual Education and Bilingualism*. 2010; (3):325–344. [PubMed: 21731899]
- Cabrera NJ, Shannon JD, West J, Brooks-Gunn J. Parental interactions with Latino infants: Variation by country of origin and English proficiency. *Child Development*. 2006; 77(5):1190–1207. [PubMed: 16999792]
- Cummins J. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*. 1979; 49(2):222–251.
- Dunn Davison M, Hammer CS. Development of 14 English grammatical morphemes in Spanish-English preschoolers. *Clinical Linguistics and Phonetics*. 2012; 26(8):728–742. [PubMed: 22774931]
- Duursma E, Romero-Contreras S, Szuber A, Proctor P, Snow C. The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics*. 2007; 28(1):171–190.
- Field, A. *Discovering statistics using SPSS*. 3rd ed.. London, UK: Sage; 2009.
- Fuligni, AJ.; Yoshikawa, H. Socioeconomic resources, parenting, and child development among immigrant families. In: Bornstein, MH.; Bradley, RH., editors. *Socioeconomic status, parenting, and child development*. Mahwah, NJ: Lawrence Erlbaum Associates Publisher; 2003. p. 107-110.

- Golberg H, Paradis J, Crago M. Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*. 2008; 29(1):41–65.
- Goldstein B. The transcription of Spanish and Spanish-influenced English. *Communication Disorders Quarterly*. 2001; 23(1):54–60.
- Goldstein B, Iglesias A. The effect of dialect on phonological analyses: Evidence from Spanish-speaking children. *American Journal of Speech-Language Pathology*. 2001; 10(3):394–406. doi: 1058-0360/01/1004-0394.
- Hammer CS, Davison MD, Lawrence FR, Miccio AW. The effect of maternal language on bilingual children's vocabulary and emergent literacy development during Head Start and kindergarten. *Scientific Studies of Reading*. 2009; 13(2):99–121. [PubMed: 23606802]
- Hammer CS, Komaroff E, Rodriguez BL, Lopez LM, Scarpino SE, Goldstein B. Predicting Spanish-English DLL children's language abilities. *Journal of Speech, Language, and Hearing Research*. 2012; 55(5):1251–1264.
- Individuals with Disabilities Education Act, 20 U.S.C. § 1400. 2004
- Kamata A. Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*. 2001; 38(1):79–93.
- Knightsoft. Study participant (Version 2) [Computer software]. Farmington, MI: Knightsoft of Michigan; 2008.
- Magis D, Béland S, Tuerlinckx F, De Boeck P. A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*. 2010; 42(3): 847–862. [PubMed: 20805607]
- Mahon EA. High stakes testing and English language learners: Questions of validity. *Bilingual Research Journal*. 2006; 30(2):479–497.
- Mathematica Policy Research. Education policy research. 2010 Retrieved from <http://www.mathematica-mpr.com/>.
- Michnowicz, J. El habla de Yucatán Spanish: A rapid and anonymous survey. In: Montreuil, JP., editor. *New Perspectives on Romance Linguistics Vol II: Phonetics, phonology, and dialectology: selected papers from the 35th Linguistic Symposium on Romance Languages (LSRL)*, Austin, Texas, February 2005. Amsterdam, NL: John Benjamins; 2005. p. 155-166.
- Motel, S.; Patten, E. The 10 largest Hispanic origin groups: Characteristics, rankings, top counties. Washington, DC: Pew Hispanic Center; 2012. Retrieved from <http://www.pewhispanic.org/2012/06/27/the-10-largest-hispanic-origin-groups-characteristics-rankings-top-counties/>
- National Center for Education Statistics (NCES). Digest of education statistics. 2009. Retrieved from <nces.ed.gov/programs/digest/>
- No Child Left Behind Act of 2001, 20 U.S.C. § 6319. 2008
- Parlakian R, Sánchez SY. Cultural influences on early language and literacy teaching practices. *Zero to Three*. 2006; 27(1):52–57. Retrieved from http://main.zerotothree.org/site/DocServer/ZTT27-1_Parlakian.pdf.
- Peña ED. Lost in translation: Methodological considerations in cross-cultural research. *Child Development*. 2007; 78(4):1255–1264. [PubMed: 17650137]
- Peña E, Halle T. Assessing preschool dual language learners: Traveling a multiforked road. *Child Development Perspectives*. 2011; 5(1):28–32.
- Place S, Hoff E. Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*. 2011; 82(6):1834–1849. [PubMed: 22004372]
- Quiroz BG, Snow CE, Zhao J. Vocabulary skills of Spanish-English bilinguals: Impact of mother-child language interactions and home language and literacy support. *International Journal of Bilingualism*. 2010; 16(4):541–565.
- Santoro M. Puerto Rican Spanish: A case of partial restructuring. *Hybrido: Arte y Literatura*. 2007; 10(9):47–57. Retrieved from <https://scholar.google.com/scholar?hl=en&sdt=0,47&q=Santoro,+M.+%282007%29+Puerto+Rican+Spanish%3A+A+case+of+partial+restructuring.+Hybrido%3A+arte+y+literatura+10,+9,+4757>.
- Schleppegrell MJ. Academic language in teaching and learning introduction to the special issue. *The Elementary School Journal*. 2012; 11(3):409–418.

- Silva-Corvalán, C. Spanish in the southwest. In: Finegan; Rickford, J., editors. *Language in the USA*. New York, NY: Cambridge; 2004. p. 205-229.
- Smith, EV., Jr; Smith, RM. *Introduction to Rasch measurement: Theory, models, and applications*. Maple Grove, MN: JAM Press; 2004.
- Solano-Flores G. Language, dialect, and register: Sociolinguistics and the estimation of measurement error in the testing of English language learners. *Teachers College Record*. 2006; 108(11):2354–2379.
- Umbel VM, Pearson BZ, Fernandez MC, Oller DK. Measuring bilingual children's receptive vocabularies. *Child Development*. 1992; 63(4):1012–1012. Retrieved from <http://www.jstor.org/stable/1131250>. [PubMed: 1505238]
- Van den Noortgate W, De Boeck P. Assessing and explaining differential item functioning using logistic mixed models. *Journal of Educational and Behavioral Statistics*. 2005; 30(4):443–464.
- Westfall, P.; Tobias, R.; Wolfinger, R. *Multiple comparisons and multiple tests using SAS*. Cary, NC: SAS Press; 2011.
- Wolfram, W. *Dialect and American English*. Englewood, NJ: Prentice Hall; 1991.
- Wolfram W, Carter P, Moriello B. Emerging Hispanic English: New dialect formation in the American south. *Journal of Sociolinguistics*. 2004; 8(3):339–358.
- Woodcock, RW.; Muñoz-Sandoval, AF.; Ruef, ML.; Alvarado, CG. *Woodcock-Muñoz language survey revised, English form*. Itasca, IL: Riverside Publishing; 2005a.
- Woodcock, RW.; Muñoz-Sandoval, AF.; Ruef, ML.; Alvarado, CG. *Woodcock-Muñoz language survey revised, Spanish form*. Itasca, IL: Riverside Publishing; 2005b.
- Zentella, A. Spanish in the northeast. In: Finegan; Rickford, J., editors. *Language in the USA*. New York, NY: Cambridge; 2004. p. 182-204.
- Zumbo B. Three generations of DIF analyses: Considering where it has been, where it is now, where it is going. *Language Assessment Quarterly*. 2007; 4(2):223–233.

FIGURE 1.
Spanish Picture Vocabulary Subtest probability of incorrect responses.

FIGURE 2.
Spanish Picture Vocabulary Subtest predicted logit of incorrect responses.

FIGURE 3.
Spanish Story Recall Subtest probability of incorrect responses.

FIGURE 4.
Spanish Story Recall Subtest predicted logit of incorrect responses.

FIGURE 5. English Picture Vocabulary Subtest probability of incorrect responses. Circled area denotes the range of items (1–14) exhibiting DIF.

FIGURE 6. English Picture Vocabulary Subtest predicted logit of incorrect responses. Circled area denotes the range of items (1–14) exhibiting DIF.

FIGURE 7.

English Story Recall Subtest probability of incorrect responses. Circled area denotes the range of items (18–27) exhibiting DIF.

FIGURE 8.

English Story Recall Subtest predicted logit of incorrect responses. Circled area denotes the range of items (18–27) exhibiting DIF.

TABLE 1

Means and Percentages of Sample Sociodemographic Characteristics of Children by Dialect ($N = 250$)

Characteristic	Mexican Dialect ($n = 105$)	Cuban Dialect ($n = 88$)	Puerto Rican Dialect ($n = 57$)
*Children's Chronological Age in months	$M = 58.32$ (4.86 years); $SD = 8.82$	$M = 59.03$ (4.91 years); $SD = 8.52$	$M = 60.49$ (5.04 years); $SD = 8.80$
Children's Gender (%)			
Female	51.40% ($n = 54$)	52.30% ($n = 46$)	63.20% ($n = 36$)
Male	48.60% ($n = 51$)	47.70% ($n = 42$)	36.80% ($n = 21$)
Attending Pre-Kindergarten	73.00% ($n = 77$)	46.43% ($n = 41$)	50.90% ($n = 29$)
Attending Kindergarten	27.00% ($n = 28$)	53.57% ($n = 47$)	49.10% ($n = 28$)
*Children's Time in U.S. Mainland ^{M-P, C-P}			
From birth	86.70% ($n = 91$)	87.50% ($n = 77$)	56.10% ($n = 32$)
Not from birth	13.30% ($n = 14$)	12.50% ($n = 11$)	43.90% ($n = 25$)
*Mothers' education ^{M-C, M-P, C-P}			
Less than high school	84.70% ($n = 89$)	44.40% ($n = 39$)	59.70% ($n = 34$)
High school diploma/GED	12.40% ($n = 13$)	26.10% ($n = 23$)	29.80% ($n = 17$)
Some college/trade school	2.90% ($n = 3$)	29.50% ($n = 26$)	10.50% ($n = 6$)
College and beyond	0.00% ($n = 0$)	0.00% ($n = 0$)	0.00% ($n = x$)
Fathers' education			
Less than high school	71.70% ($n = 75$)	10.80% ($n = 10$)	34.60% ($n = 20$)
High school diploma/GED	17.20% ($n = 18$)	39.80% ($n = 35$)	38.20% ($n = 22$)
Some college/trade school	8.10% ($n = 9$)	26.50% ($n = 23$)	23.60% ($n = 13$)
College and beyond	3.00% ($n = 3$)	22.90% ($n = 20$)	3.60% ($n = 2$)
Father lives at home	94.30% ($n = 99$)	80.50% ($n = 71$)	80.70% ($n = 46$)
Father does not live at home	5.70% ($n = 6$)	19.50% ($n = 17$)	19.30% ($n = 11$)

Note. All data are percentages except for children's chronological age, which is presented with the M = mean and the SD = standard deviation.

* The variable was used as a covariate in the DIF analysis;

M-C A significant difference ($p < .05$) between Mexican and Cuban; M-P a significant difference ($p < .05$) between Mexican and Puerto Rican; C-P a significant difference ($p < .05$) between Cuban and Puerto Rican.

TABLE 2

Percentages of Language Exposure by Dialect ($N = 250$)

Dialect Subgroup	Exposure	*Mothers to Children <i>M-C, M-P, C-P</i>	*Fathers to Children <i>M-P</i>	*Teachers to Children <i>M-C, M-P, C-P</i>
Mexican $n = 105$	More or all Spanish	88.57% ($n = 93$)	78.10% ($n = 82$)	16.19% ($n = 17$)
	Equal Spanish & English	9.52% ($n = 10$)	14.29% ($n = 15$)	40.00% ($n = 42$)
	More or all English	1.90% ($n = 2$)	7.62% ($n = 8$)	43.81% ($n = 46$)
Cuban $n = 88$	More or all Spanish	67.05% ($n = 59$)	69.32% ($n = 61$)	7.95% ($n = 7$)
	Equal Spanish & English	19.31% ($n = 17$)	15.91% ($n = 14$)	26.14% ($n = 23$)
	More or all English	13.64% ($n = 12$)	14.77% ($n = 13$)	65.91% ($n = 58$)
Puerto Rican $n = 57$	More or all Spanish	57.89% ($n = 33$)	71.93% ($n = 41$)	0.00% ($n = 0$)
	Equal Spanish & English	14.04% ($n = 8$)	7.02% ($n = 4$)	7.02% ($n = 4$)
	More or all English	28.07% ($n = 16$)	21.05% ($n = 12$)	92.98% ($n = 53$)

Note.

* The variable was used as a covariate in the DIF analysis;

M-C_A significant difference ($p < .05$) between Mexican and Cuban; M-P_A significant difference ($p < .05$) between Mexican and Puerto Rican; C-P_A significant difference ($p < .05$) between Cuban and Puerto Rican.

TABLE 3

Percentages of language usage by dialect (N = 250)

Dialect Subgroup	Usage	*Children to Mothers <i>M-C, M-P, C-P</i>	*Children to Fathers <i>M-C, M-P, C-P</i>	*Children to Teachers <i>M-C, M-P, C-P</i>
Mexican	More or all Spanish	82.86% (<i>n</i> = 87)	77.14% (<i>n</i> = 81)	53.33% (<i>n</i> = 56)
<i>n</i> = 105	Equal Spanish & English	13.33% (<i>n</i> = 14)	13.33% (<i>n</i> = 14)	16.19% (<i>n</i> = 17)
	More or all English	3.81% (<i>n</i> = 4)	9.52% (<i>n</i> = 10)	30.48% (<i>n</i> = 32)
Cuban	More or all Spanish	62.50% (<i>n</i> = 55)	57.95% (<i>n</i> = 51)	28.41% (<i>n</i> = 25)
<i>n</i> = 88	Equal Spanish & English	14.77% (<i>n</i> = 13)	17.05% (<i>n</i> = 15)	20.45% (<i>n</i> = 18)
	More or all English	22.73% (<i>n</i> = 20)	25.00% (<i>n</i> = 22)	51.14% (<i>n</i> = 45)
Puerto Rican	More or all Spanish	31.58% (<i>n</i> = 18)	42.11% (<i>n</i> = 24)	10.53% (<i>n</i> = 6)
<i>n</i> = 57	Equal Spanish & English	17.54% (<i>n</i> = 10)	12.28% (<i>n</i> = 7)	7.02% (<i>n</i> = 4)
	More or all English	50.88% (<i>n</i> = 29)	45.61% (<i>n</i> = 26)	82.46% (<i>n</i> = 47)

Note.

* The variable was used as a covariate in the DIF analysis;

M-C A significant difference ($p < .05$) between Mexican and Cuban; M-P A significant difference ($p < .05$) between Mexican and Puerto Rican; C-P A significant difference ($p < .05$) between Cuban and Puerto Rican.

TABLE 4

Means and Standard Deviations of Spanish and English WMLS-R Raw and Standard Scores

<i>WMLS-R Subtest</i>	<i>Raw Scores</i>		<i>Standard Scores</i>	
	M	SD	M	SD
Spanish Picture Vocabulary				
Mexican	16.57	6.04	83.55	20.22
Cuban	13.78	6.09	73.67	21.95
Puerto Rican	9.89	5.70	56.32	27.18
Spanish Story Recall				
Mexican	7.15	4.21	74.29	26.62
Cuban	7.52	3.89	76.42	23.30
Puerto Rican	6.23	3.67	66.54	30.77
English Picture Vocabulary				
Mexican	9.80	7.61	49.44	32.31
Cuban	17.29	8.01	78.99	30.88
Puerto Rican	16.02	5.97	75.24	23.07
English Story Recall				
Mexican	3.98	3.88	62.24	39.95
Cuban	7.69	4.44	88.90	23.99
Puerto Rican	7.38	4.15	88.83	18.19

Note. Standard scores have a mean of 100 and a standard deviation of 15.

TABLE 5

Significant DIF Items on English Picture Vocabulary Subtest

English Picture Vocabulary				
DIF Items	Higher Probability of Incorrect Response	Logit Difference	Bonferroni Adjusted p-value	Test Item
1	M	−2.23	.000	ball
2	M	−2.16	.000	mouth
3	M	−2.08	.000	window
4	M	−2.00	.000	balloons
5	M	−1.93	.000	cat
6	M	−1.85	.000	chair
7	M	−1.78	.000	house
8	M	−1.70	.000	apple
9	M	−1.63	.000	sock
10	M	−1.55	.000	glasses
11	M	−1.47	.003	boat
12	M	−1.40	.007	ice cream cone
13	M	−1.32	.017	cow
14	M	−1.25	.037	toothbrush

Note. M = Mexican.

TABLE 6

Significant DIF Items on English Story Recall

English Story Recall				
DIF Items	Group More Difficult For	Logit Difference	Bonferroni Adjusted <i>p</i> -value	Test Item (Passage 5)
18	M	1.60	.045	/The baby bear/
19	M	1.70	.032	/scrambled up the tree/
20	M	1.79	.023	/He looked down/
21	M	1.89	.018	/ proudly /
22	M	1.98	.014	/at his brother/
23	M	2.10	.011	/who was far below/
24	M	2.17	.001	/Suddenly he realized that/
25	M	2.27	.001	/climbing down was harder than climbing up/
26	M/C	2.37	.001	/He gave out a loud, piercing cry/
27	M/C	2.46	.001	/for his mother/

Note. M = Mexican; C = Cuban. Bold items must be stated verbatim.